

PENGELOMPOKAN KABUPATEN/KOTA DI JAWA TENGAH MENURUT KUALITAS UDARA AMBIEN MENGGUNAKAN ANALISIS K-HARMONIC MEAN CLUSTER (Studi Kasus: Kualitas Udara Ambien pada Kawasan Pemukiman di Jawa Tengah Tahun 2015)

Sandra Wijaya, Moch. Abdul Mukid, Dwi Ispriyanti
Staf Pengajar Departemen Statistika FSM UNDIP, Semarang
mamukid@live.undip.ac.id

ABSTRACT

Ambient air is the open air which is vital to the survival of every living being on earth. The activities of people have directly impacted the ambient air quality which can be seen by the result of the ambient air quality measuring. The measuring can be done in residential areas toward some substances which parameters have been measured: Nitrogen Dioxide (NO₂), Carbon Monoxide (CO), Sulfur Dioxide (SO₂), Oxide (O₃), Hydrocarbon (HC), dan Total Suspended Particulate /Dust (TSP). To figure out the characteristic similarity between 35 Districts in Central Java, we need to group them based on the substances' parameter to help the government run the ambient air pollutant reducing program in each clusters. The method used is K-Harmonic Means (KHM) Cluster. KHM is an algorithm invented to complete the K-Means. In this method, the centroid is updated by put the harmonic average from all points into each centroid clusters. There would be two, three, four, five, or six clusters. We need to use the ratio of variability within clusters and ratio of variability between clusters to get a relevant count of how many clusters are. The final result shows that five clusters is the relevant amount of cluster. In the first cluster there is 1 District with the highest levels of TSP, O₃, SO₂ and NO₂. In the second cluster there are 2 Districts with the highest level of CO and the level of NO₂ as the lowest. In the third cluster there are 6 Districts with the highest CO level and the lowest O₃ level. In the fourth cluster there are 21 Districts with the highest HC level, but the levels of CO, TSP, and SO₂ in this cluster are lower than in the third cluster. In the fifth cluster there is 1 District with the highest level of HC but the CO, SO₂, and TSP levels are the lowest in this cluster.

Keywords: ambient air, K-Harmonic Means Cluster, K-Means, centroid, cluster

1. Latar Belakang

Udara adalah campuran gas yang terdapat pada lapisan yang mengelilingi bumi, yang mana komposisi campuran gas tersebut tidak selalu konstan (Fardiaz, 1992). Udara dibedakan menjadi udara emisi dan udara ambien. Udara emisi yaitu udara yang dikeluarkan oleh sumber emisi seperti knalpot kendaraan bermotor dan cerobong gas buang industri. Sedangkan udara ambien adalah udara bebas di permukaan bumi pada lapisan troposfir yang berada di dalam wilayah yurisdiksi Republik Indonesia yang dibutuhkan dan mempengaruhi kesehatan manusia, makhluk hidup dan unsur lingkungan hidup lainnya (PP No.41 Tahun 1999). Udara ambien inilah yang biasa dimanfaatkan untuk kegiatan sehari-hari seperti mengeringkan pakaian, menggerakkan kincir angin, membantu proses penyerbukan, dan lain-lain. Namun, udara

ambien dapat dicemari oleh udara emisi apabila pengelolaan lingkungan tidak baik.

Pada era globalisasi saat ini, kualitas udara ambien semakin menurun akibat peningkatan sumber pencemar udara oleh manusia seperti dari kendaraan bermotor, kegiatan industri, pembakaran sampah dan lain-lain. Untuk mendapatkan udara ambien yang baik perlu dilakukan pengendalian pencemaran udara yang dapat dilakukan salah satunya dengan memantau atau mengukur baku mutu udara ambien.

Baku mutu udara ambien adalah ukuran batas atau kadar zat, energi, dan komponen yang ada /atau yang seharusnya ada dan unsur pencemar yang ditenggang keberadaannya dalam udara ambien (PLH NO. 12 tahun 2010). Pengukuran kualitas udara ambien di suatu daerah dapat dilakukan di kawasan pemukiman, kawasan industri, dan kawasan padat lalu lintas di mana kawasan-kawasan tersebut banyak terjadi

kegiatan manusia. Pengukuran kualitas udara ambien dilakukan terhadap zat-zat yang sudah ditentukan parameternya di antaranya adalah SO₂ (Sulfur Dioksida), NO₂ (Nitrogen Dioksida), CO (Karbon Monoksida), O₃ (Oksida), TSP (Total Suspended Particulate /Debu), dan HC (Hidrokarbon). Apabila zat tersebut tidak melebihi baku mutu udara maka dikatakan zat tersebut dalam kondisi baik, sedangkan zat yang melebihi baku mutu udara dikatakan zat dalam kondisi buruk dan berpotensi mencemari udara ambien.

Menurut Badan Pusat Statistika pada tahun 2014 pencemaran udara ambien di Jawa Tengah mengalami peningkatan dari tahun sebelumnya, yang mana disebabkan oleh peningkatan sumber pencemar udara yang terdiri dari sumber tetap (tungku industri, pembangkit tenaga, kilang minyak, dan lain-lain), sumber bergerak (kendaraan darat, kapal, turbin gas tetap, kereta api, dan lain-lain), dan sumber pembuangan.

Penerapan analisis cluster dapat digunakan untuk mengelompokkan 35 Kabupaten/Kota Provinsi Jawa Tengah berdasarkan parameter zat yang mempengaruhi udara ambien. Pengelompokan dilakukan untuk mengidentifikasi kesamaan karakteristik antara 35 kabupaten/kota sehingga pemerintah dapat menjalankan program-program yang sesuai pada tiap cluster guna mengurangi pencemaran udara ambien.

Analisis cluster adalah teknik pengelompokan objek-objek yang mempunyai kesamaan kedalam suatu kelompok yang sama (Han dan Kamber, 2006). Menurut Simamora (2005), terdapat beberapa metode yang dapat digunakan dalam analisis cluster salah satunya metode partitioning clustering. Salah satu metode partitioning clustering yang sangat populer adalah K-Means cluster karena algoritmanya sangat mudah diimplementasikan, namun K-Means memiliki beberapa kelemahan, di antaranya hasil cluster sensitif terhadap penentuan awal (inisialisasi) pusat cluster. Hasil yang sering berbeda dari proses clustering pada data yang sama disebabkan oleh inisialisasi titik pusat (centroid) yang berbeda.

Untuk mengatasi masalah tersebut, pada tahun 1999, Zhang, Hsu, dan Dayal memperkenalkan sebuah metode K-Harmonic Mean Cluster. K-Harmonic Means Cluster merupakan salah satu

metode cluster yang bertujuan untuk mengelompokkan objek berdasarkan karakteristik yang dimilikinya dengan penyempurnaan secara iteratif berdasarkan letak centroid dari masing-masing cluster. Pada algoritma ini, setiap titik data dicari jaraknya ke semua centroid. Rata-rata harmonik yang sensitif terhadap fakta adanya dua atau lebih centroid yang berada di dekat suatu titik data, secara natural algoritma ini akan menukar satu atau lebih centroid ke area dimana terdapat titik data yang tidak memiliki centroid di dekatnya. Sehingga rata-rata harmonik yang digunakan dalam metode K-Harmonic Means Cluster mampu mengurangi permasalahan tersebut.

Berdasarkan uraian di atas penulis tertarik untuk melakukan pengelompokan Kabupaten/Kota di Jawa Tengah menurut kualitas udara ambien pada kawasan pemukiman di Jawa Tengah dengan menggunakan analisis K-Harmonic Means Cluster.

2. Tinjauan Pustaka

2.1. Udara Ambien

Menurut PP NO. 41 Tahun 1999 tentang pengendalian pencemaran udara, udara ambien adalah udara bebas di permukaan bumi pada lapisan troposfir yang berada di dalam wilayah yurisdiksi Republik Indonesia yang dibutuhkan dan mempengaruhi kesehatan manusia, makhluk hidup, dan unsur lingkungan hidup lainnya. Jika perubahan komposisi udara alami melebihi konsentrasi tertentu yang menyebabkan udara ambien tidak dapat memenuhi fungsinya, maka dikatakan udara tersebut telah tercemar (PLH No. 12 tahun 2010).

Udara merupakan campuran gas yang terdapat pada lapisan yang mengelilingi bumi yang mana komposisi campuran gas tersebut tidak konsisten (Fardiaz, 1992). Komposisi udara bersih 78,09% Nitrogen, 20,94% Oksigen, 0,93% Argon, 0,0032% CO₂, dan sisanya unsur lainnya. Komposisi udara tersebut sangat ideal untuk kehidupan baik manusia, tumbuhan maupun hewan. Unsur-unsur berbahaya yang masuk kedalam atmosfer dapat berupa Karbon Monoksida (CO), Nitrogen Dioksida (NO₂), Hidrokarbon (HC), Sulfur Dioksida (SO₂), dan partikel lainnya. Unsur-unsur berbahaya tersebut

dikatakan sebagai bahan pencemar apabila unsur tersebut melebihi baku mutu udara ambien.

Menurut Peraturan Pemerintah No.14 tahun 1999 tentang pengendalian udara, baku mutu udara ambien adalah ukuran batas atau kadar zat, energi dan/atau komponen yang ada atau yang

seharusnya ada dan/atau unsur pencemar yang ditenggang keberadaannya dalam udara ambien. Setiap daerah memiliki baku mutu udara ambien sendiri, adapun baku mutu udara ambien menurut SK Gubernur Jawa Tengah No.8 tahun 2001 adalah sebagai berikut:

Tabel 1 Baku Mutu Udara Ambien menurut SK Gubernur Jawa Tengah No.8 Tahun 2001

Parameter	Waktu Pengukuran	Baku Mutu	Peralatan
NO ₂ (Nitrogen Dioksida)	1 jam	316 µg/Nm ³	Spektrofotometer
	24 jam	150 µg/Nm ³	
	1 tahun	100 µg/Nm ³	
CO (Karbon Monoksida)	1 jam	15.000 µg/Nm ³	NDIR Analyzer
	24 jam	10.000 µg/Nm ³	
SO ₂ (Sulfur Dioksida)	1 jam	632 µg/Nm ³	Spektrofotometer
	24 jam	365 µg/Nm ³	
	1 tahun	60 µg/Nm ³	
O ₃ (Oksida)	1 jam	200 µg/Nm ³	Spektrofotometer
	1 tahun	50 µg/Nm ³	
HC (Hidrokarbon)	3 jam	160 µg/Nm ³	Gas Chromatografi
TSP (Total Suspended Partikulate /Debu)	24 jam	230 µg/Nm ³	Hi-Vol
	1 tahun	90 µg/Nm ³	

Udara ambien yang tercemar memiliki kualitas udara yang buruk. Udara yang melebihi baku mutu dapat merusak lingkungan sekitarnya dan berpotensi mengganggu kesehatan masyarakat sekitarnya. Pencemaran udara ambien dapat menimbulkan efek negatif terhadap lingkungan seperti: hujan asam, penipisan lapisan ozon, dan pemanasan global.

Perlindungan mutu udara ambien merupakan upaya yang dilakukan agar udara ambien dapat memenuhi fungsi sebagaimana mestinya. Pengukuran kualitas udara ambien adalah salah satu cara untuk melindungi baku mutu udara ambien.

2.2. Rata-rata Harmonik (Harmonic Average)

Menurut Sudjana (1991), rata-rata harmonik untuk sekelompok data didefinisikan sebagai kebalikan rata-rata hitung dari kebalikan nilai-nilai data tersebut. Jika rata-rata harmonik dinyatakan dengan H dan data dengan x_1, x_2, \dots, x_n , maka:

$$H = \frac{n}{\sum_{i=1}^n \left(\frac{1}{x_i}\right)}$$

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}$$

dengan:

H : rata-rata harmonik

n : banyak data

x_i : nilai datum ke- i

2.3. K-Harmonic Means Cluster

Menurut Prasetyo (2014) K-Harmonic Means (KHM) merupakan salah satu metode clustering yang menggunakan prosedur partisi dengan menggunakan rata-rata harmonik (harmonic average) jarak dari setiap titik data ke centroid sebagai komponen dalam fungsi tujuan. KHM merupakan salah satu metode clustering berbasis titik pusat yang diperkenalkan oleh Zhang, Hsu dan Dayal pada tahun 1999. KHM dikembangkan untuk menangani masalah utama dalam K-Means yang hasil clustering-nya sangat sensitif terhadap inisialisasi data yang dijadikan centroid awal. Hasil yang sering berbeda (local optima) dari proses clustering disebabkan oleh inisialisasi centroid yang berbeda. Pada KHM cluster dibentuk dengan penyempurnaan secara iteratif berdasarkan letak titik pusat dari masing-masing cluster.

Nilai fungsi tujuan pada KHM dihasilkan dengan mencari total rata-rata harmonik dari seluruh titik data terhadap jarak antara masing-masing titik data ke seluruh titik pusat cluster yang ada. Hal ini berbeda dengan K-means dimana fungsi tujuan diperoleh dari total jarak keseluruhan data ke titik

pusat cluster-nya. Fungsi tujuan dari KHM adalah sebagai berikut:

$$KHM(X, C) = \sum_{i=1}^n \frac{K}{\sum_{j=1}^K \frac{1}{\|x_i - c_j\|^2}}$$

dimana:

K = jumlah cluster

$X = \{x_1, \dots, x_n\}$ adalah himpunan amatan

x_i = nilai datum ke- i

n = banyak data

$C = \{c_1, \dots, c_n\}$ adalah titik pusat dari cluster

c_j = pusat cluster

Rata-rata harmonik sangat sensitif terhadap fakta adanya dua atau lebih titik pusat yang saling berdekatan. Algoritma ini secara natural akan menukar satu atau lebih titik pusat ke area dimana terdapat suatu titik data yang tidak memiliki titik pusat didekatnya. Sehingga semakin baik hasil cluster-nya dan nilai fungsi tujuan akan semakin kecil (Widiartha, 2011). Adapun algoritma KHM menurut Yang, et al (2009) adalah sebagai berikut:

1. Tentukan banyaknya cluster yang akan dibuat (K)
2. Tentukan centroid awal dari masing-masing cluster
3. Hitung jarak dari semua titik objek ke setiap centroid dengan menggunakan jarak Euclid
4. Hitung fungsi tujuan
5. Untuk setiap objek (x_i), hitung keanggotaan $m(c_j|x_i)$ pada setiap pusat c_j berdasarkan:

$$m(c_j|x_i) = \frac{\|x_i - c_j\|^{-p-2}}{\sum_{j=1}^K \|x_i - c_j\|^{-p-2}}$$

dengan $p \geq 2$ dan $i = 1, 2, \dots, n$

6. Untuk setiap objek (x_i), hitung bobot $w(x_i)$ berdasarkan:

$$w(x_i) = \frac{\sum_{j=1}^K \|x_i - c_j\|^{-p-2}}{(\sum_{j=1}^K \|x_i - c_j\|^{-p-2})^2}$$

7. Untuk setiap pusat c_j , hitung ulang lokasinya terhadap semua objek (x_i) berdasarkan keanggotaan dan bobotnya:

$$c_j = \frac{\sum_{i=1}^n m(c_j|x_i)w(x_i)x_i}{\sum_{i=1}^n m(c_j|x_i)w(x_i)}$$

8. Ulangi langkah 3 sampai 7 hingga proses iterasi mencapai konvergen dimana $|KHM_n - KHM_{n-1}| < \epsilon$
9. Masukkan objek (x_i) ke cluster j dengan nilai $m(c_j|x_i)$ yang terbesar.

Dalam penelitian ini, jumlah cluster yang dibuat sebanyak dua, tiga, empat, lima dan enam cluster,

dimana setiap kelompok memiliki karakteristik yang sama.

2.4. Pemilihan Ukuran Jarak

Menurut Supranto (2004), karena tujuan melakukan cluster adalah mengelompokkan objek yang mirip ke dalam cluster yang sama, maka beberapa ukuran jarak diperlukan untuk mengetahui seberapa mirip atau berbeda objek-objek tersebut. Menurut Simamora (2006) beberapa jarak yang digunakan dalam analisis cluster yaitu :

- Jarak Euclid adalah akar dari jumlah selisih antar objek yang dikuadratkan. Jarak Euclid dengan p variabel antara objek ke-m dan ke-n didefinisikan sebagai berikut:

$$d_{m,n} = \sqrt{\sum_{j=1}^p (x_{mj} - x_{nj})^2}$$

dengan

$d_{m,n}$: jarak euclid

$x_{mj} - x_{nj}$: selisih objek ke-m dan ke-n pada variabel j

Jarak Euclid antara kedua titik adalah panjang dari sisi terpanjang dari segitiga siku-siku (Hair et al. 2006).

- Kuadrat Jarak Euclid adalah jumlah selisih kuadrat jarak tanpa mengambil akar kuadratnya. Dengan tidak mengambil akar kuadratnya kembali, proses perhitungan menjadi lebih cepat. Jarak Kuadrat Euclid dengan p variabel antara objek ke-m dan ke-n didefinisikan sebagai berikut:

$$d_{m,n}^2 = \sum_{j=1}^p (x_{mj} - x_{nj})^2$$

dimana:

$d_{m,n}^2$: kuadrat jarak euclid

$x_{mj} - x_{nj}$: selisih objek ke-m dan ke-n pada variabel j (j = 1, 2, 3, ..., p)

- Jarak Cityblock atau jarak Manhattan adalah jumlah selisih absolut tiap objek. Jarak Cityblock atau jarak Manhattan dengan p variabel antara objek ke-m dan ke-n didefinisikan sebagai berikut:

$$d_{m,n} = \left[\sum_{j=1}^p |x_{mj} - x_{nj}| \right]$$

- Jarak Chebychev merupakan jarak antara dua objek dengan selisih nilai absolut maksimum pada setiap variabel.

Pada tahap pemilihan ukuran jarak, dipilih jarak Euclid sebagai ukuran kemiripan dari suatu objek dengan ukuran jarak yang pendek. Semakin pendek

ukuran jarak antar objek maka dapat dikatakan kedua objek memiliki kemiripan. Dan sebaliknya, apabila jarak semakin panjang dapat dikatakan kedua objek tidak memiliki kemiripan.

2.5. Penentuan Jumlah Cluster

Penentuan cluster yang relevan diperlukan analisis lebih lanjut yaitu analisis terhadap variabilitas dalam cluster (trace W) dan variabilitas antar cluster (trace B). Cluster yang relevan adalah cluster yang memiliki nilai trace W terkecil dan nilai trace B terbesar. Trace W merupakan jumlah dari diagonal utama matriks W sedangkan trace B merupakan jumlah dari diagonal utama matriks B (Dillon dan Goldstein, 1984). Menurut Johnson dan Wichern (2007), untuk menghitung matriks W dan B adalah sebagai berikut:

$$W = \sum_{j=1}^K \sum_{i=1}^{n_j} (\mathbf{x}_{ji} - \mathbf{c}_j)(\mathbf{x}_{ji} - \mathbf{c}_j)' \quad (19)$$

$$B = \sum_{j=1}^K n_j (\mathbf{c}_j - \bar{\mathbf{c}})(\mathbf{c}_j - \bar{\mathbf{c}})'$$

dengan :

n_j : jumlah observasi dalam cluster

\mathbf{c}_j : centroid cluster ke-j

$\bar{\mathbf{c}}$: rata-rata centroid cluster

\mathbf{x}_{ji} : observasi dalam cluster

3. Data dan Metode

Data yang digunakan adalah data sekunder yang diperoleh dari Badan Lingkungan Hidup, Provinsi Jawa Tengah. Data tersebut merupakan data hasil pengukuran kualitas udara ambien pada kawasan pemukiman di kabupaten/kota di Jawa Tengah pada tahun 2015. Parameter zat yang diperhatikan meliputi SO₂, NO₂, CO, O₃, TSP, dan HC yang diukur menggunakan satuan µg/Nm³.

Pengolahan data pada penelitian ini yaitu menggunakan software Microsoft Excel 2013, SPSS 16, dan Matlab R2014a. Adapun Langkah-langkah yang dilakukan untuk menganalisis data adalah :

- Melakukan uji asumsi multikolinieritas
- Melakukan Analisis cluster dengan tahapan sebagai berikut:
 - Tentukan banyaknya cluster yang akan dibuat (k)
 - Tentukan centroid awal dari masing-masing cluster
 - Hitung jarak dari semua titik objek ke setiap centroid dengan menggunakan jarak Euclid
 - Hitung fungsi tujuan dengan formula:

$$KHM(X, C) = \sum_{i=1}^n \frac{K}{\sum_{j=1}^K \frac{1}{\|\mathbf{x}_i - \mathbf{c}_j\|^2}}$$

5. Untuk setiap objek (x_i), hitung keanggotaan $m(c_j|x_i)$ pada setiap pusat c_j berdasarkan:

$$m(c_j|x_i) = \frac{\|x_i - c_j\|^{-p-2}}{\sum_{j=1}^K \|x_i - c_j\|^{-p-2}}$$

6. Untuk setiap objek (x_i), hitung bobot $w(x_i)$ berdasarkan:

$$w(x_i) = \frac{\sum_{j=1}^K \|x_i - c_j\|^{-p-2}}{(\sum_{j=1}^K \|x_i - c_j\|^{-p})^2}$$

7. Untuk setiap pusat c_j , hitung ulang lokasinya terhadap semua objek (x_i) berdasarkan keanggotaan dan bobotnya:

$$c_j = \frac{\sum_{i=1}^n m(c_j|x_i)w(x_i)x_i}{\sum_{i=1}^n m(c_j|x_i)w(x_i)}$$

8. Ulangi langkah 3 sampai 7, hingga nilai fungsi tujuan tidak berubah secara signifikan
9. Masukkan objek (x_i) ke cluster j dengan nilai $m(c_j|x_i)$ yang terbesar.
3. Tentukan cluster yang relevan
4. Interpretasi hasil cluster.

4. Hasil dan Pembahasan

4.1. Analisis Deskriptif

Data yang digunakan merupakan data hasil pengukuran kualitas udara ambien pada kawasan pemukiman di Kabupaten/Kota di Provinsi Jawa Tengah Tahun 2015 yang diperoleh dari Badan Lingkungan Hidup (BLH) Provinsi Jawa Tengah Bidang Pengendalian Pencemaran Lingkungan. Dari data tersebut ingin diketahui pengelompokkan Kabupaten/Kota di Jawa Tengah menurut kualitas udara pada kawasan pemukiman dengan parameter zat diantaranya SO_2 , NO_2 , CO , O_3 , TSP, dan HC. Jika zat tersebut tidak melebihi baku mutu udara maka dikatakan zat tersebut dalam kondisi baik, sedangkan zat yang melebihi baku mutu udara dikatakan zat dalam kondisi buruk dan berpotensi mencemari udara ambien. Untuk gambaran secara umum kualitas udara ambien pada kawasan pemukiman dapat dilihat pada Tabel 3.

Tabel 3 Gambaran Umum Kualitas Udara Ambien pada Kawasan Pemukiman

Parameter Zat	Maksimum ($\mu g/Nm^3$)	Minimum ($\mu g/Nm^3$)	Rata-Rata ($\mu g/Nm^3$)
NO_2	54,59	1,59	12,19
CO	6524,00	11,45	1086,22
SO_2	42,52	12,10	16,55
HC	2838,00	340,30	973,15
TSP	7192,14	24,34	389,60
O_3	53,32	1,94	15,52

Dari Tabel 3 dapat diketahui kadar NO_2 tertinggi

terdapat di Kab. Salatiga dengan hasil ukur 54,59 $\mu g/Nm^3$ dan hasil ukur terendah sebesar 1,59 $\mu g/Nm^3$ terdapat pada Kab. Jepara. Hasil ukur untuk CO tertinggi dengan hasil ukur 6524,00 $\mu g/Nm^3$ terdapat pada Kab. Pemalang sedangkan hasil ukur terendah yaitu 11,45 $\mu g/Nm^3$ terdapat pada 5 Kabupaten/Kota yaitu Kab. Semarang, Kab. Demak, Kab. Kudus, Kab. Rembang, dan Kab. Batang.

Hasil ukur untuk SO_2 tertinggi yaitu 42,52 $\mu g/Nm^3$ terdapat pada Kab. Wonosobo sedangkan hasil ukur terkecil yaitu 12,10 $\mu g/Nm^3$ terdapat pada Kab. Kendal dan Kab. Brebes. Untuk hasil ukur HC tertinggi terdapat pada Kota Semarang yaitu sebesar 2838,00 $\mu g/Nm^3$ sedangkan hasil ukur terendah sebesar 340,30 $\mu g/Nm^3$ terdapat pada Kab. Tegal.

Kadar TSP tertinggi terdapat di Kab. Kebumen dengan hasil ukur 7192,14 $\mu g/Nm^3$ dan hasil ukur terendah sebesar 24,34 $\mu g/Nm^3$ terdapat di Kab. Blora. Dan untuk hasil ukur O_3 terbesar yaitu 53,32 $\mu g/Nm^3$ terdapat pada Kab. Sragen, terkecil yaitu 1,94 $\mu g/Nm^3$ pada Kab. Jepara.

4.2. Asumsi Analisis Cluster

Terdapat dua uji asumsi dalam analisis cluster yaitu sampel mewakili populasi (sample representatif) dan tidak terjadi multikolinieritas.

1. Sampel mewakili populasi

Untuk melakukan sebuah penelitian sampel yang digunakan harus mewakili populasi yang ada. Tidak ada ketentuan untuk jumlah sampel yang representatif, namun tetaplah diperlukan sejumlah sampel yang cukup besar agar proses clustering dapat dilakukan dengan benar. Karena data yang digunakan pada penelitian ini merupakan populasi seluruh Kab/Kota di Jawa Tengah, maka asumsi sampel mewakili populasi (sample representatif) terpenuhi.

2. Asumsi Multikolinieritas

Untuk memeriksa data yang digunakan mengalami multikolinieritas atau tidak, dapat diketahui dengan cara menghitung nilai VIF. Nilai VIF dapat diketahui dengan persamaan (10) dengan R_j^2

merupakan koefisien determinasi dan indeks j menyatakan variabel independen ke j .

Koefisien determinasi atau R_j^2 dari setiap variabel diperoleh dengan menjadikan variabel yang ingin diketahui nilai koefisien determinasinya sebagai variabel dependen dan variabel sisanya sebagai variabel independen. Koefisien determinasi untuk NO_2 , CO , SO_2 , HC, TSP, O_3 , dapat lihat pada Lampiran 2. Untuk nilai VIF dapat dilihat pada

Tabel 4.

Tabel 4 Nilai VIF dari Enam Variabel		
Variabel	R2	VIF
NO2	0,078	1,0846
CO	0,128	1,1468
SO2	0,242	1,3193
HC	0,175	1,2121
TSP	0,365	1,5748
O3	0,235	1,3072

Tabel 4 menunjukkan bahwa keenam variabel penelitian di atas tidak terdapat nilai VIF yang memiliki nilai lebih dari 10. Oleh karena itu, dapat disimpulkan tidak terdapat pengaruh multikolinieritas antar variabel.

5. Proses Clustering

Dalam penelitian ini dilakukan pengelompokkan Kabupaten/Kota di Jawa Tengah menurut kualitas udara ambien pada kawasan pemukiman menggunakan analisis K-Harmonic Mean Cluster (KHM). Proses pengelompokkan menggunakan KHM ini dicoba dengan berbagai ukuran cluster. Untuk proses clustering, penelitian ini menggunakan perangkat lunak Matlab R2014a. Perintah yang digunakan untuk proses clustering dapat dilihat pada Lampiran 3 sampai Lampiran 7 yang mana dengan menetapkan nilai awal sebagai berikut:

1. Banyaknya cluster (K) = 2, 3, ..., 6
2. Batas toleransi (ϵ) = 10-6
3. Fungsi tujuan awal = 0

Proses clustering akan berhenti jika iterasi mencapai konvergen dimana $|KHM_n - KHM_{n-1}| < \epsilon$.

Berdasarkan perhitungan di atas, maka dapat dibuat tabel antara jumlah cluster, nilai trace W dan trace B yang terlihat seperti Tabel 10.

Tabel 10 Nilai Trace W dan Trace B untuk K = 2, 3, ..., 6

Jumlah cluster	tr(W)	tr(B)
2	78240613,946	136792988,241
3	69039551,389	165409079,616
4	58879477,916	111443530,735
5	15701256,082	239963551,203
6	55842152,355	126710706,048

Pada Tabel 5 dapat dilihat bahwa jumlah cluster yang terbaik adalah 5 cluster, karena tr(W) pada 5

cluster memiliki nilai terkecil di antara cluster yang lain dan nilai tr(B) pada 5 cluster memiliki nilai terbesar di antara cluster yang lain. Sehingga, cluster lima memiliki variabilitas dalam cluster yang minimum dan variabilitas antar cluster yang maksimum.

6. Interpretasi dan Pembuatan Profil

Menurut Simamora (2005), menginterpretasi dan pembuatan profil dapat dilakukan dengan rata-rata nilai objek yang terdapat dalam cluster pada tiap variabel (yang dinamakan centroid). Berdasarkan pembahasan sebelumnya diperoleh bahwa cluster terbaik adalah 5 cluster, dengan nilai centroid setiap cluster dapat dilihat pada Tabel 6 sedangkan untuk anggota dari setiap cluster dapat dilihat pada Lampiran 11.

Tabel 11 Centroid Setiap Cluster

Zat	cluster 1	cluster 2	cluster 3	cluster 4	cluster 5
NO2	19,151	4,415	11,876	11,906	12,111
CO	3558,610	5730,520	2159,515	334,821	261,545
SO2	36,046	14,319	17,649	16,525	12,859
HC	548,682	612,546	727,328	761,976	1785,248
TSP	7172,364	139,564	256,039	161,460	123,928
O3	49,162	20,892	10,653	12,047	20,722

Informasi yang bisa diperoleh dari Tabel 11 adalah:

1. Cluster satu merupakan cluster yang memiliki kadar TSP, O3, SO2, dan NO2 tertinggi dari cluster yang lainnya, namun kadar NO2 merupakan kadar terendah dalam cluster satu. Cluster satu terdiri dari satu Kabupaten/Kota yaitu Kabupaten Kebumen.
2. Cluster dua merupakan cluster yang memiliki kadar CO tertinggi dan memiliki kadar NO2 terendah dari cluster yang lainnya maupun dalam cluster dua itu sendiri. Cluster dua terdiri dari 2 Kabupaten/Kota yaitu Kabupaten Jepara dan Kabupaten Pemalang
3. Cluster tiga memiliki kadar CO tertinggi dari cluster yang lainnya dan memiliki kadar NO2, SO2, HC dan TSP yang lebih tinggi dari cluster dua. Pada cluster tiga kadar O3 lebih rendah dari cluster yang lainnya. Cluster tiga terdiri dari 6 Kabupaten/Kota yaitu Kota Semarang, Kota Tegal, Kabupaten Tegal, Kota Surakarta, Kabupaten Purbalingga, Kabupaten Wonosobo
4. Cluster empat merupakan cluster yang memiliki kadar HC tertinggi dan kadar NO2 terendah dalam clusternya. Kadar CO, TSP, dan SO2 pada cluster empat lebih rendah dibandingkan dengan cluster tiga. Cluster

- empat terdiri dari 21 Kabupaten/Kota yaitu Kabupaten Semarang, Kota Salatiga, Kabupaten Kendal, Kabupaten Demak, Kabupaten Kudus, Kabupaten Grobogan, Kabupaten Rembang, Kota Pekalongan, Kabupaten Batang, Kabupaten Brebes, Kabupaten Boyolali, Kabupaten Klaten, Kabupaten Karanganyer, Kabupaten Wonogiri, Kabupaten Magelang, Kota Magelang, Kabupaten Banyumas, Kabupaten Banjarnegara, Kabupaten Cilacap, Kabupaten Pekalongan, dan Kabupaten Temanggung
5. Cluster lima memiliki kadar HC tertinggi dari cluster yang lainnya dan memiliki kadar CO, SO₂, dan TSP terendah dari cluster yang lainnya. Cluster lima terdiri dari 5 Kabupaten/Kota yaitu Kabupaten Pati, Kabupaten Sukoharjo, Kabupaten Sragen, Kabupaten Purworejo, dan Kabupaten Blora.
7. Kesimpulan
Kabupaten/ Kota di Jawa Tengah dapat dikelompokkan menjadi lima cluster berdasarkan metode K Harmonic Means. Hal ini dapat dilihat dari nilai variabilitas dalam cluster (trace W) dan variabilitas antar cluster (trace B) dengan nilai trace W terkecil dan nilai trace B terbesar. Dengan menggunakan cluster ini, pemerintah Provinsi Jawa Tengah dapat menentukan kebijakan-kebijakan terkait dengan perbaikan mutu udara di setiap kabupaten secara efektif.
- Daftar Pustaka
- [BSN]. Badan Standardisasi Nasional. 2005. SNI 19-7119.2-2005 Tentang Udara Ambien – Bagian 2: Cara Uji Kadar Nitrogen Dioksida (NO₂) dengan metoda Griess Saltzman Menggunakan Spektrofotometer. Jakarta: BSN.
- [BSN]. Badan Standardisasi Nasional. 2005. SNI 19-7119.3-2005 Tentang Udara Ambien – Bagian 3: Cara Uji Partikel Tersuspensi Total (TSP) dengan Menggunakan Peralatan High Volume Air Samplar (HVAS) dengan metoda Gravimetri. Jakarta: BSN.
- [BSN]. Badan Standardisasi Nasional. 2005. SNI 19-7119.7-2005 Tentang Udara Ambien – Bagian 7: Cara Uji Kadar Sulfur Dioksida (SO₂) dengan metoda Pararosalin Menggunakan Spektrofotometer. Jakarta: BSN.
- [BSN]. Badan Standardisasi Nasional. 2005. SNI 19-7119.8-2005 Tentang Udara Ambien – Bagian 8: Cara Uji Kadar Oksidan (O₃) dengan Metode Neural Buffer Kalium iodida (NKBI) Menggunakan spektrofotometer. Jakarta: BSN.
- [BSN]. Badan Standardisasi Nasional. 2011. SNI 19-7119.10-2011 Tentang Udara Ambien - Bagian 10: Cara Uji Konsentrasi Kabon Monoksida (CO) Menggunakan Metode Nondispersive Infrared (NDIR) . Jakarta: BSN.
- [BSN]. Badan Standardisasi Nasional. 2009. SNI 19-7119.13-2009 Tentang Udara Ambien – Bagian 13: Cara Uji Kadar Hidrokarbon (HC) menggunakan hydrocarbon analyzer dengan detektor ionisasi nyala (Flame Ionization Detector/FID). Jakarta: BSN.
- Dillon, W.R. dan Goldstein, M. 1984. Multivariate Analysis Methods and Application. New York: John Wiley & Sons, Inc.
- Fardiaz, S. 1992. Polusi Air dan Udara. Yogyakarta: Kasinius.
- Gozali, I. 2006. Aplikasi Multivariat dengan Program SPSS. Semarang: BP Universitas Diponegoro.
- Han, Jiawai dan Kember, M. 2006. Data Mining Concepts and Techniques. secon Edition. San Francisco: Elsevier.
- Hair, J.F.Jr. et al. 2006. Multivariat Data Analysis 6th Edition. New Jersey: Pearson Prentice Hill, Inc.
- Johnson, R.A. dan Wichern, D.W. 2007. Applied Multivariate Statistical Analysis Sixth Edition. New Jersey: Prentice Hill, Inc.
- Kementrian Negara Lingkungan Hidup. 2010. Peraturan Menti Lingkungan Hidup No. 12 Tahun 2010 Tentang Pelaksanaan Pengendalian Pencemaran Udara di Daerah. Jakarta: Sekretaris Negara.
- Machfudhoh, S., Wahyuningsih, N. 2013. Analisis Cluster Kabupaten/Kota Berdasarkan Pertumbuhan Ekonomi Jawa Timur. Jurnal Sains dan Seni POMITS, Vol.2, Nomor.1, hal 2337-3520.
- Pemerintah Daerah Provinsi Jawa Tengah. 2001. Surat Keputusan Gubernur Jawa Tengah Nomor 8 Tahun 2001 Tentang Baku Mutu Udara Ambien di Provinsi Jawa Tengah. Sekretariat Daerah, Jawa Tengah

Prasetyo, E. 2014. Mengolah Data Menjadi Informasi Menggunakan Matlab. Yogyakarta: Andi.

BIBLIOGRAPHY Rawlings, J. O. et all. 1998. Applied Regression Analysis: A Research Tool. New York: Springer.

Republik Indonesia. 1999. Peraturan Pemerintah No. 41 Tahun 1999 Tentang Pencemaran Udara. Lembaran Negara RI Tahun 1999, No. 86. Jakarta: Sekretariat Negara

Rencher, A. C. 2002. Methods of Multivariate Analysis Second Edition. New York: John Wiley and Sons, Inc.

Simamora, B. 2005. Analisis Multivariat Pemasaran. Jakarta: PT gramedia pustaka umum.

Sudjana. 1991. Statistik Untuk Ekonomi dan Niaga. Bandung: Tarsito.

Supranto, J. 2004. Analisis Multivariat: Arti dan Interpretasi. Jakarta: PT Rineka Cipta.

Widiartha, I. M. 2011. Studi Komparasi Metode Klasterisasi Data K-means dan K-Harmonic Means. Jurnal Ilmu Komputer, Vol.4, Nomor.1.

Yang, F., Sun, T., Zhang, C. 2009. An Efficient Hybrid Data Clustering Method Based on K-Harmonic Means and Particle Swarm Optimization. Expert System With Applications, 9847-9852.

Zhang, B., Hsu, M., dan Dayal, U. 1999. K-Harmonic Means- A Data Clustering Algorithm. HP Laboratories Palo ALto.

_____. BPS. 2014. Data dan informasi Lingkungan Hidup Jawa Tengah Tahun 2014